

Auswertung von Drucksondierungen dargestellt am Beispiel der Data Mining - Methoden

Problemstellung und Zielsetzung

Bodenkenngrößen, die als Ausgangsbasis für die Festlegung charakteristischer Werte nach DIN V ENV 1997-1:1996-04, 2.4.3 verwendet werden können, sollen aus Versuchsergebnissen abgeleitet werden. Drucksondierungen als indirekte Aufschlüsse des Bodens liefern hingegen keine expliziten Werte der geotechnischen Parametern. Entsprechend DIN 4094-1:2002 müssen daher quantitative Zusammenhänge zwischen dem Sondierergebnis und einer geotechnischen Kenngröße für die jeweilige Bodenart innerhalb definierter Gültigkeitsgrenzen in nachprüfbarer Form statistisch ermittelt oder deterministisch festgelegt werden. Klassische statistische Methoden sind aber oft kompliziert in der Anwendung und liefern nicht selten auch nur schwer durchschaubare Ergebnisse.

Lösungsansatz

Data Mining, als eine neue rechnergestützte Methoden der Datenanalyse, soll auf das Problem der Auswertung von Drucksondierungen angepaßt werden.

Data Mining definiert sich als einen Prozeß, Muster und Strukturzusammenhänge in Daten zu erkennen und aus großen Datenmengen Regeln zu erzeugen.

Data Mining benutzt verschiedene Verfahren:

- Assoziationsregeln

Assoziationsregeln besagen, dass mit dem Eintreten bestimmter Ereignisse X häufig auch die Ereignisse Y auftreten. Die Zuverlässigkeit (confidence) der Regel gibt an, in wieviel Prozent der Fälle die Implikation tatsächlich zutrifft. Der Wert support besagt, in wie vielen Fällen die Regel überhaupt anwendbar ist.

$X \Rightarrow Y$ (confidence, support).

- Klassifizierungsmethoden

Eine Klassifizierungsregel versucht, den Wert einer abhängigen Zielvariablen (die Klasse) aus den Werten von bekannten Variablen vorauszusagen. Klassifizierungsregeln werden dann benötigt, wenn fehlende Informationen aus bekannten Eigenschaften von Objekten abgeleitet werden sollen. Eine der Klassifizierungsmethoden ist die Klassifizierung anhand der Entscheidungsbäume.

- Bayes Klassifikation

Diese Methode basiert auf der Regel der bedingten Wahrscheinlichkeit von Bayes. Sie besagt, dass wenn man eine Hypothese H und ein zu dieser Hypothese passendes Ereignis E hat, dann für die Wahrscheinlichkeit von H gegeben durch das Ereignis E

$P(H|E) = P(E|H)P(H) / P(E)$ gilt.

- Clustering

Es werden Klassen durch geschicktes Gruppieren der Objekte in Cluster gebildet. Eine möglichst starke Trennung der Klassen untereinander wird dadurch erreicht, dass die Objekte innerhalb einer Klasse einander sehr ähnlich sind.

- Genetische Algorithmen

Diese Algorithmen dienen der Suche des globalen Extremwertes einer Funktion. Folgende Schritte werden dabei ausgeführt:

- Selektion;
- genetische Operationen (Kreuzung und Mutation).

Als Chromosom wird ein Baum verwendet, dessen Knoten Elemente der vom Benutzer vorgegebenen Basisfunktionen (Summierung, Division usw.) darstellen.

Bisherige Erkenntnisse

Es wurden mehrere Berechnungen zur Untersuchung der Korrelationen zwischen den Bodeneigenschaften und aus der Literatur entnommenen überregionalen Daten aus Drucksondierungen mit Hilfe des Data Mining-Pakets im SQL-Server2000 von Microsoft durchgeführt. Unter anderem wurde für bindige Böden der Einfluß des effektiven Spitzendrucks und des auftretenden Porenwasserüberdrucks, ausgedrückt durch die Werte $(q_t - u_1) / \sigma_{v0}'$ und $B_q = \Delta u_2 / (q_t - \sigma_{v0})$, auf den Überkonsolidierungsgrad OCR untersucht, wobei q_t der korrigierte Spitzenwiderstand, σ_{v0} die totale Vertikalspannung, σ_{v0}' die effektive Vertikalspannung und B_q das Porenwasserdruckverhältnis sind. Der Wert u_1 stellt den Porenwasserdruck auf den konischen Teil des Kegels der Sonde dar. Die Kenngröße Δu_2 entspricht der Differenz zwischen dem Eindringporenwasserdruck auf den zylindrischen Teil des Kegels und dem Porenwasserdruck u_0 , der im Boden vorhanden ist, bevor die Drucksondierung durchgeführt wird.

Es wurde festgestellt, dass bei starker Überkonsolidierung ($OCR > 1,5$) der Einfluß des Parameters $(q_t - u_1) / \sigma_{v0}'$ auf die OCR-Werte bedeutend höher ist als von B_q , während im Fall niedrigerer Werte des Überkonsolidierungsgrades die Kenngröße B_q die entscheidende Rolle bei der Ermittlung von OCR spielt.

Mit Hilfe des Data Mining-Programms See5 (www.rulequest.com) wurden für nichtbindige Böden die Zusammenhänge zwischen der bezogenen Lagerungsdichte des Bodens, der mittleren Spannung an der Sondierungsstelle und dem Spitzendruck der Sonde untersucht. Die Anwendung der in diesem Programm implementierten Klassifizierungsmethoden liefert einfach zu interpretierende Ergebnisse, die mit den aus der Literatur bekannten Gesetzmäßigkeiten übereinstimmen.

Beispiel

Klassifizierung eines nichtbindigen Boden nach dem Spitzendruckwiderstand (q_c - Wert):

Definition der Klassen:

Klasse	q_c [MPa]
1	< 15
2	15 - 30
3	30 - 47

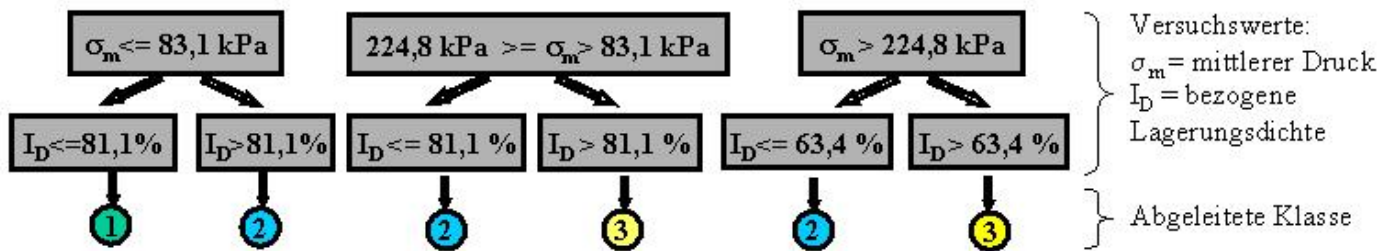


Abbildung 1: Klassifizierung nach Entscheidungsbaum:

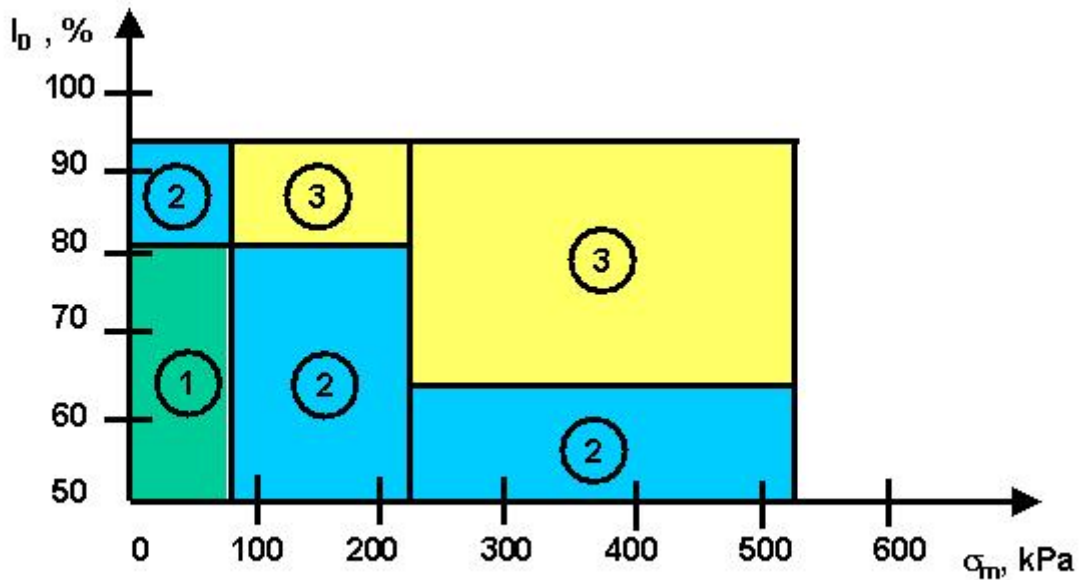


Abbildung 2: Ergebnis der Datenunterteilung in Klassen entsprechend dem Entscheidungsbaum

Ausblick

Data Mining – Verfahren sollen derart auf die Auswertung von Drucksondierungsergebnissen und anderen indirekten Erkundungsmethoden angepaßt werden, dass sie eine optimale Suche nach den Gesetzmäßigkeiten und Zusammenhängen zwischen den Meßergebnissen und den Bodenparametern ermöglichen und darüber hinaus die Streuung der Kenngrößen (Rauschen in Daten) berücksichtigen. Es werden die am Institut vorhandenen Software - Pakete, die unterschiedliche Verfahren von Data Mining repräsentieren, auf ihre Eignung für die Auswertung von Drucksondierungen untersucht.

Mit Hilfe von Data Mining sollen die Regeln für eine Zuordnung der Böden in definierte Klassen sowie für eine Ermittlung ihrer mechanischen Eigenschaften aufgrund der Ergebnisse indirekter Bodenuntersuchungsmethoden entwickelt und die Gültigkeitsbereiche der gefundenen Zusammenhänge und Regeln untersucht werden.

Literatur

- Ziegler, M.; Prokhorova, A.: Verbesserung der Interpretation von Sondierungsergebnissen mit Hilfe von Data Mining-Methoden. In: Vorträge der Baugrundtagung 2004 in Leipzig / Deutsche Gesellschaft für Geotechnik e.V. Essen, Verlag Glückauf, 2004. S. 213–220.
- Prokhorova, A.: Anwendung der Data Mining - Methoden für Prognosen in der Geotechnik. In: Ziegler, M. (Hrsg.): Stochastische Prozesse in der Geotechnik. Aachen, Mainz, 2003. S. 65-79. (Schriftenreihe Geotechnik im Bauwesen ; 1)
- Prokhorova, A.; Ziegler, M.: Better understanding of field measurements by means of data mining. In: Natau, O.; Fecker, E.; Pimentel, E. (Eds.): Geotechnical measurements and modelling. Lisse (u.a.): Balkema, 2003. S. 291-296.
- Prokhorova, A.; Ziegler, M.: Anwendung von Data mining in der Geotechnik. In: 2. Siegener Kolloquium Erkundung und Felduntersuchungen in der Geotechnik. Siegen: Universitätsverl., 2003. S. 1-21.

Bearbeiter: Dipl.-Ing. Alla Prokhorova
Lehrstuhl für Geotechnik im Bauwesen der RWTH-Aachen
Tel. : 0241 8025254
E-mail: prokhorova@geotechnik.rwth-aachen.de